A new mixture model for the estimation of credit card exposure at default

Citation for published version:
Leow, M & Crook, J 2016, 'A new mixture model for the estimation of credit card exposure at default', *European Journal of Operational Research*, vol. 249, no. 2, pp. 487-497.
https://doi.org/10.1016/j.ejor.2015.10.001

Digital Object Identifier (DOI):
10.1016/j.ejor.2015.10.001

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
European Journal of Operational Research

A New Mixture Model for the Estimation of Credit Card Exposure at Default

Mindy Leow[a,b] & Jonathan Crook[c]

[a]Credit Research Centre, University of Edinburgh Business School, 29 Buccleuch Place, Edinburgh EH8 9JS, Scotland, United Kingdom
[b]Contact details: mindy.leow@ed.ac.uk
[c]Corresponding author: +44 131 650 3802; j.crook@ed.ac.uk

A New Mixture Model for the Estimation of Credit Card Exposure at Default

Abstract

Using a large portfolio of historical observations on defaulted loans, we estimate Exposure at Default at the level of the obligor by estimating the outstanding balance of an account, not only at the time of default, but at any time over the entire loan period. We theorize that the outstanding balance on a credit card account at any time during the loan is a function of the spending by the borrower and is also subject to the credit limit imposed by the card issuer. The predicted value is modelled as a weighted average of the estimated balance and limit, with weights depending on how likely the borrower is to have a balance greater than the limit. The weights are estimated using a discrete-time repeated events survival model to predict the probability of an account having a balance greater than its limit. The expected balance and expected limit are estimated using two panel models with random effects. We are able to get predictions which, overall, are more accurate for outstanding balance, not only at the time of default, but at any time over the entire default loan period, than any other particular technique in the literature.

Keywords: risk management, forecasting, panel models, survival models, macroeconomic variables, time-varying covariates

1. Introduction

Predictions of Exposure At Default (EAD) are useful to banks for at least two reasons. First, the Basel Accords define expected loss as the product of Probability of Default (PD), Loss Given Default (LGD) and EAD, so predictions of EAD are needed to compute Regulatory Capital. Second, predictions of EAD are needed for the prediction of Economic Capital that a bank believes it needs to protect its depositors in the event of severe unexpected events. Since the credit crisis of 2008, there has been increased awareness of the models for these components, and in particular, for retail loans. However, these have been mainly focused on PD and LGD models, and how they should and can be improved (see Thomas (2010) for a review). The analysis and modelling of EAD at account level has so far been relatively neglected. For loans with fixed loan amounts over fixed terms and pre-agreed monthly repayment amounts, it is possible to estimate at least a reasonable range for EAD should the loan be expected to default in the following time horizon, e.g. in the next 12 months. However, in the case of revolving loans, the subject of this paper, i.e. loans with no fixed loan amount or term, debtors are given a line of credit, with a credit limit up to which they can draw upon at any time (as long as they have not gone into default). This could make it difficult for financial institutions to predict account level outstanding balance should an account go into default, especially if accounts deteriorate into default quickly and draw heavily on the card just before default.

Another issue associated with the analysis and modelling of EAD is the measurement of EAD. EAD is similar to LGD in that its value is only of interest in the event default occurs (although its value still needs to be estimated for the calculation and preparation of economic capital). However, unlike LGD, where loss is predicted to be at some time point after default, EAD is known the very instant the account goes into default. Therefore, although default-time variables could be used in the modelling of LGD, they cannot be used for EAD models. As such, practitioners and the literature create various indicators to be estimated instead of EAD, taking into account the current balance and available limit. Unfortunately, each method has limitations.

Our aim is to propose a new method to predict EAD for each loan in a portfolio and to demonstrate its accuracy by comparisons with methods currently in use and in the literature. Unlike conventional cross section methods, our proposed approach exploits the panel nature of a typical credit card dataset to model the values of balance and limit over time in a way that allows extrapolation from the time of prediction to the time of default. To evaluate our model, we use a large portfolio of defaulted loans and their historical observations, to directly estimate EAD at the level of the obligor by estimating the outstanding balance of an account, not only at the time of default, but *at any time* over the entire loan period, up to the time of default.

Our methodology has several advantages over current methods. First for revolving credit loans, balance typically approaches the limit as an account moves over time towards default. We exploit this observation, to the extent that it is true, and the observation that modelling an account's limit at each time in its history can be done more accurately than the balance to more accurately predict the balance at default (that is EAD) than if this information is not used. Second we avoid several of the problems associated with current methods of modelling EAD which we describe in section 2, for example the considerable sensitivity to very small values of a denominator. Third by using panel models we can more accurately include the effects of macroeconomic variables and so enable EAD estimates to be fixed as in a down-turn scenario than cross sectional models. Further our method yields predictions of balance at any time in an account's history and a bank would benefit from such predictions to estimate expected future interest income and so a component of~~n~~ expected profit from an account.

The development and validation of the new Mixture model contributes to the literature in two ways. First, this is the first paper to predict the outstanding balance for defaulted loans at any time during the life of a revolving loan. Second, we incorporate macroeconomic variables into the model and so provide a framework suitable for stress testing later. The rest of this paper is structured as follows. Section 2 reviews the literature and Section 3 explains the model. In Section 4, we illustrate the use of the method and compare its performance to methods in the literature. Section 5 shows an empirical application and Section 6 concludes.

2. EAD in the literature

Only a few papers have examined EAD and usually for corporate loans (see e.g. Araten and Jacobs (2001), Jacobs Jr. (2008), Jiménez and Mencía (2009), Jiménez et al. (2009), Yang and Tkachenko (2012) and Barkova and Pathasarathy (2013)). Few consider account level models, and they do not model EAD directly (e.g. see Taplin et al. (2007), Risk Management Association (2004)). Instead, they typically model the Loan Equivalent Exposure (LEQ) Factor, the Credit Conversion Factor (CCF) or the Exposure At Default Factor (EADF)[1], and then transform them back to an estimate of EAD (a more comprehensive review can be found in Moral (2006)). Thus, Jacobs Jr. (2008), using corporate data and a GLM modelling framework, models all three factors. Barakova and Parthasarrathy (2013) model four ratios using four algorithms applied to corporate level variables for large syndicated loans over 2007 to 2009. Yang and Tkachenko (2012) model EADF using eight account level variables and compare seven estimators applied to 500 commercial borrowers. The closest to our work is Qi (2009), who used unsecured credit card data, to model LEQ by looking at the level of credit drawn at one year before default. No macroeconomic variables were included in the above models. All come to the conclusion that EAD plays an important part in the calculation of the provision of capital and should be more carefully incorporated into risk and loss calculations.

To define these terms, we adopt the definitions as in Jacobs Jr. (2008), Qi (2009) Barakova and Parthasarathy (2013) and Yang and Tkachenko (2012). In terms of nomenclature from here on, outstanding balance of account *i* at duration time τ is represented by $B_{i\tau}$, and limit of account *i* at duration time τ is represented by $L_{i\tau}$. We also construct a binary variable $d_{i\tau}$ that takes on the value 1 if account *i* defaults at time τ and d_i that takes on the value 1 if account *i* defaults at some time in the future. To simplify the notation, the subscript *i* representing account *i* is dropped for the equations in this sub-section. The three variables EADF_D, CCF_D and LEQ_D are defined in Table 1.

[1] Note that LEQ, CCF and EADF are not universally defined. Basel II refers to a Credit Conversion Factor, "CCF", but does not define it except to state that it is a factor of any further undrawn limit (see BASEL COMMITTEE ON BANKING SUPERVISION 2004. International Convergence of Capital Measurement and Capital Standards: A Revised Framework., Paragraph 316, 474-478), so it is not clear that there is a standard industry practice towards EAD modelling.

Table 1: Definition of EAD measures in use and in the literature

Variable		Explanation
$\mathrm{EADF}_D = \left\{ \dfrac{B_D}{L_{D-l}} \right.$	for $L_{D-l} \neq 0$	Ratio of the balance at default time *D*, over the limit at observation time $D-l$; limit is usually the limit at the time of application and is known once account is opened
$\mathrm{CCF}_D = \begin{cases} \dfrac{B_D}{B_{D-l}} \\ 0 \end{cases}$	if $B_{D-l} \neq 0$ if $B_{D-l} = 0$	Ratio of the balance at default time *D* over the balance at some observation time $D-l$; this tries to get better predictions for balance by taking into account the outstanding balance of an account at some observation time before default.
$\mathrm{LEQ}_D = \begin{cases} \dfrac{B_D - B_{D-l}}{L_{D-l} - B_{D-l}} \\ 0 \end{cases}$	if $L_{D-l} \neq B_{D-l}$ if $L_{D-l} = B_{D-l}$	A more sophisticated prediction for balance by not only taking into account balance at some observation time before default, $D-l$, but also the undrawn limit at that time, i.e. the remaining amount of credit the debtor is able to draw upon.

However, modelling EAD in terms of these ratios involves a number of difficulties, some of which are rehearsed by Jacobs Jr. (2008) and Qi (2009). In the case of EADF_D, although we expect its value to range between 0 and 1, it is possible and quite common to see outstanding balances greater than the assigned limits, perhaps due to accumulated interest or banks allowing borrowers to go over their limits, giving values much greater than 1. This makes the choice of distribution slightly more challenging. A further problem noted by Qi (2009) is that as an account moves towards default and its balance increases, lenders may respond differently between accounts; in some cases increasing the limit, in others reducing the limit. This may introduce unexplained heterogeneity in a cross sectional model of EADF_D.

Considering CCF_D, it is possible that the outstanding balance at the selected observation time happens to be £0, or even negative (the account is in credit), which would give $\mathrm{CCF}_D = 0$, and this raises the issue of the treatment of these accounts. It is also possible that some of these accounts then deteriorate quickly into delinquency and default. Also, should the account have a very low balance during observation time and defaults with a large balance, CCF_D could become an extremely large value, causing difficulties with data analysis and model estimation. Although on the one hand, it is likely that accounts that go into default have large balances on their account prior to default (for example, debtors who default due to behavioural issues), it is also possible that accounts go from a low or zero balance to default within a short period of time (for example, debtors who default due to unexpected circumstances), which could then imply a different set of predictors for each group. From the point of prediction, a value of 0 for CCF_D does not make any sense as this would mean a prediction of £0 for balance at some time in the future, and possibly at default.

Our method does not suffer from the theoretical inability to deal with zero or negative values of balance or the difficulty in modelling a dependent variable which is composed of a ratio where its value is very sensitive to different values of the denominator. In our approach we use panel data that incorporates unexplained heterogeneity unlike cross sectional models that have been used for the above ratios.

The different values that the LEQ_D can take could arise due to a number of different situations and which would give different implications. Should the account have zero undrawn limit, i.e. outstanding balance equal to limit, at the time of observation, we get an LEQ_D value of 0. This is a group of debtors who have used their maximum available limit and are likely to default, but would be difficult to include and handle in the modelling because the LEQ_D value computed does not have the same implications as the other LEQ_D values computed for when balance and limit are not equal.

The majority of accounts would have a positive LEQ_D, which could be due to one of two situations: (a) when balance at default is greater than balance at observation, and balance at observation is below the credit limit at observation, which would be the most common progression into default; or (b) when balance at observation is greater than balance at default, and balance at observation is already greater than the limit at observation. The latter would represent debtors who are actually recovering from a large balance (and where perhaps extending the credit without putting the account into default might give lower loss). Although these two groups of debtors would have LEQ_D in the same range, we expect their characteristics and circumstances to be quite different. It is also possible to have negative LEQ_D again in different situations[2]. The possible range of LEQ_D, coupled with the fact that different types of borrowers and circumstances could give LEQ_D in the same range, would make it difficult to estimate and model LEQ_D.

One weakness of several of the above methods is that according to how they are defined, these variables could become unstable[3] if the denominator is very small, so some restrictions have to be imposed on the range of values. Qi (2009) included only accounts at default time where undrawn limit is greater than 50 USD; Jacobs Jr. (2008) restricted the values of LEQ to between 0 and 1 and replaced outliers with the maximum and minimum values of his selected range. In his CCF model, he restricted the range of CCF to between the 1 and 99 percentiles, and replaced outliers with these maximum and minimum values. Both authors effectively ignored accounts that go from up-to-date to default suddenly or within a short time period, but this was the only way to get plausible results. Barakova and Parthasarathy (2013) winsorise LEQ and CCF at the 99th percentile. Yang and Tkachenko (2012) capped EADF at 1 and floored it at 0. Taplin et al. (2007) did not attempt to estimate LEQ (referred to as "CCF" in their paper) as they would have to exclude about 50% of their observations. They proposed regression models that estimate EAD as a function of balance and limit, but did not give any

[2] These are: (a) when balance at observation is larger than limit at observation and balance at default is larger than balance at observation, which would represent debtors who are spiralling further into debt and default; or (b) when balance at observation is larger than balance at default, but both are below the limit at observation. Again, we have two groups of debtors with negative LEQ_D values but where they have arrived via different circumstances.

[3] These variables could have large volatility over short periods of time, most likely coinciding with the period just before default occurs as balance on accounts go from small to large in a short period of time.

indication of covariates used or any performance measures. Note also that predictive results from most papers in the literature that used these dependent variables have generally been poor.

3. The new Mixture model

We propose the prediction of outstanding balance using a Mixture model. The random variable, balance of account *i* at duration time τ could be above, equal to or below the account limit. The expected balance for account *i* at time τ is therefore given in Equation 1:

$$\begin{aligned} E(B_{i\tau} \mid d_i = 1) = & \left(P(B_{i\tau} > L_{i\tau} \mid d_i = 1) \times E(B_{i\tau} \mid B_{i\tau} > L_{i\tau}, d_i = 1)\right) \\ & + \left(P(B_{i\tau} = L_{i\tau} \mid d_i = 1) \times E(B_{i\tau} \mid B_{i\tau} = L_{i\tau}, d_i = 1)\right). \\ & + \left(P(B_{i\tau} < L_{i\tau} \mid d_i = 1) \times E(B_{i\tau} \mid B_{i\tau} < L_{i\tau}, d_i = 1)\right) \end{aligned} \quad (1)$$

Typically, as an account moves towards default, the balance increases towards and may exceed the limit. Often, borrowers stop increasing the balance when it reaches the limit. We exploit this occurrence in our method. Balance is less systematically governed by a model than is the limit, which is the result of a model. Instead of modelling $E(B_{i\tau} \mid B_{i\tau} > L_{i\tau}, d_i = 1)$ directly, we assume, *as an approximation*, that such accounts have an expected balance equal to their limit and replace Equation 1 by Equation 2.

$$\begin{aligned} E(B_{i\tau} \mid d_i = 1) = & \left(P(B_{i\tau} \geq L_{i\tau} \mid d_i = 1) \times E(L_{i\tau} \mid B_{i\tau} \geq L_{i\tau}, d_i = 1)\right) \\ & + \left(P(B_{i\tau} < L_{i\tau} \mid d_i = 1) \times E(B_{i\tau} \mid B_{i\tau} < L_{i\tau}, d_i = 1)\right) \ . \end{aligned} \quad (2)$$

We therefore propose the parameterisation of three models. First, a model of the probability that the outstanding balance of an account is larger than the credit limit, conditional on default; second, a model to predict the outstanding balance, conditional on default; and third, a model to predict the credit limit conditional on default, where the parameters to predict balance and limit are allowed to differ.

There are cases where the limit may not increase and may even decrease as balance increases (see Qi (2009) for a good discussion of this). But our method is robust to this situation in that for such cases the survival model would be expected to predict a higher probability that in the next month the predicted balance will exceed the limit and so the weight on predicted balance in that month will be correspondingly lower and the weight on the predicted limit correspondingly higher, as Equation 2 shows.

From a training dataset based on only default accounts, i.e. accounts that eventually go into default, we propose the estimation of the probability that the outstanding balance at any duration time τ is equal to or greater than the limit at duration time τ. This is done by defining the event 'overstretched', $S_{i\tau}$, for account *i* at time τ which takes the value 1 if outstanding balance is greater than the limit at time τ; 0 otherwise, given in Equation 3:

$$S_{i\tau} = \begin{cases} 1 & \text{if } B_{i\tau} \geq L_{i\tau} \\ 0 & \text{otherwise} \end{cases} . \tag{3}$$

Given this definition, it is possible for an account to experience the event more than once (at different times of the loan), so a discrete-time repeated events survival model, given in Equation 4, is estimated.

$$\log\left(\frac{P(S_{i\tau})}{1-P(S_{i\tau})}\right) = \nu + \alpha(\tau) + \beta_1 X_i + \beta_2 Y_{i,\tau-l} + \beta_3 Z_{\tau-l}, \tag{4}$$

where ν is the intercept term; $\alpha(\tau)$ is a function of time since the last event; X_i are account-dependent, time-independent covariates, i.e. application variables; $Y_{i,\tau-l}$ are account-dependent, time-dependent covariates, lagged *l* months, i.e. behavioural variables; $Z_{\tau-l}$ are account-independent, time-dependent covariates, lagged *l* months, i.e. macroeconomic variables; and $\beta_1, \beta_2, \beta_3$ are unknown vectors of parameters to be estimated.

To predict either balance or limit, we propose the estimation of two sub-models using two separate training datasets (where we use entire histories of the accounts in each training set). The datasets consist of accounts that at some time in their history defaulted as shown in Figure 1. The training dataset is segmented according to whether accounts ever had balance exceeding limit (but not necessarily in default) at any point in the loan, or accounts that never had balance exceeding limit throughout the life of the loan. The subset consisting of accounts (represented by subscript *a*) where balance exceeded credit limit at some point during the loan is the limit training set, and used to estimate the limit at time τ, conditional on default. By structuring a sample in this way, our method involves parameterising the distribution of $L_{a\tau}$ given $B_{a\tau} \geq L_{a\tau}$ and given default. The other subset consisting of accounts (represented by subscript *b*) where balance never exceeded limit throughout the observation time of the loan is the balance training set, and used to estimate the balance at time τ. Hence, our method parameterises the $B_{b\tau}$ given the $B_{b\tau} < L_{b\tau}$ distribution. By segmenting the accounts in this way, we use the full history of each account in the estimation of either balance or limit as it changes over time and over the course of the loan period. This methodology, as well as the training and test sets created (details in the next section), is represented in Figure 1.

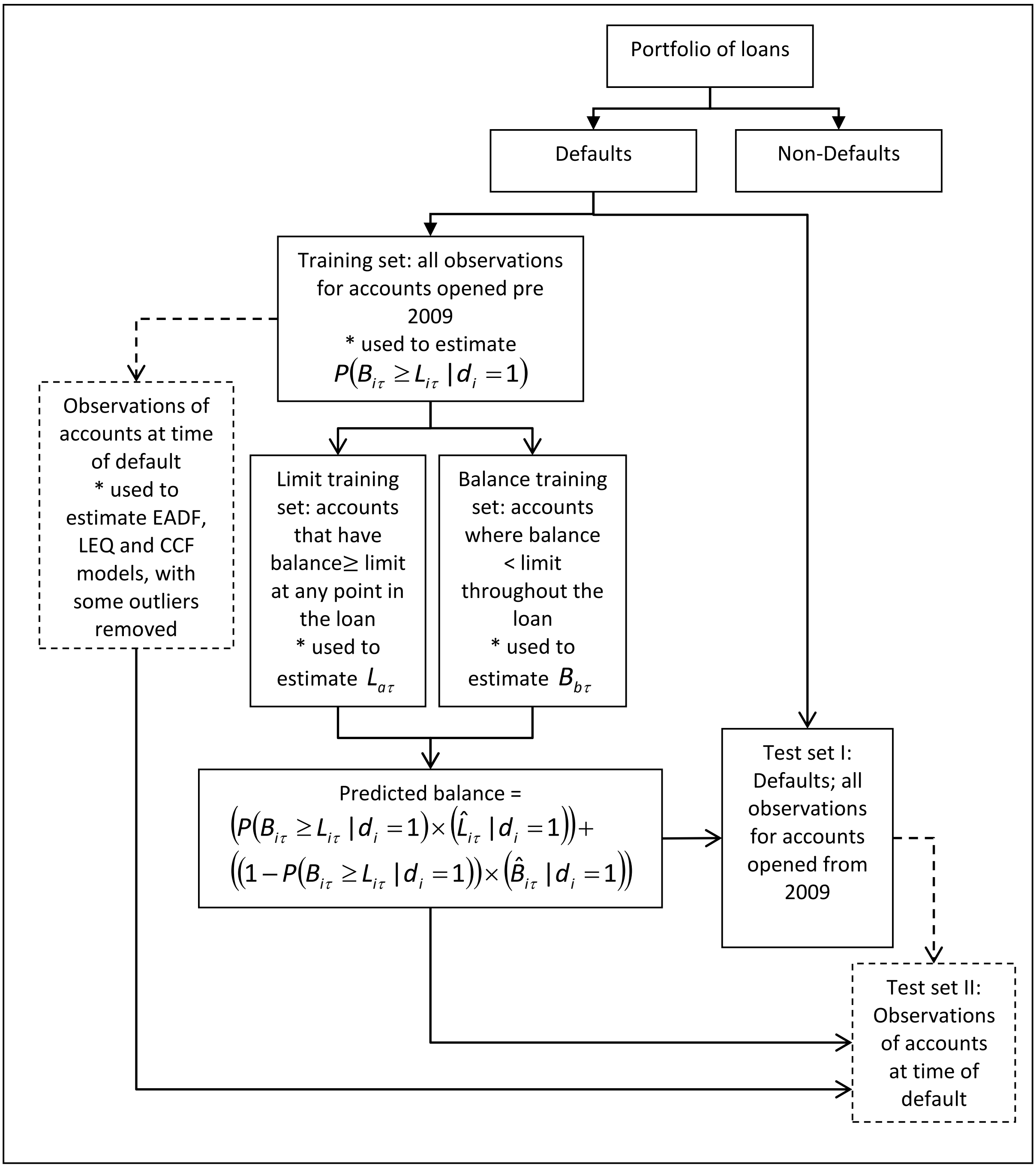


Figure 1: Flowchart of methodology and training and test set splits, where dotted lines represent subsets of the test and training sets that only consist of observations at default time.

The limit, $L_{a\tau}$, and balance, $B_{b\tau}$, for accounts *a* and *b,* respectively, at time τ could be estimated using panel models with random effects given in Equations 5 and 6 (see Cameron and Trivedi (2005), Gujarati (2003) and Verbeek (2004) for details).

$$\left[\hat{L}_{a\tau} \mid d_a = 1\right] = \mu^L + \gamma_1^L X_a + \gamma_2^L Y_{a,\tau-l} + \gamma_3^L Z_{\tau-l} + \alpha_a + \varepsilon_{a\tau} \qquad (5)$$

$$\left[\hat{B}_{b\tau} \mid d_b = 1\right] = \mu^B + \gamma_1^B X_b + \gamma_2^B Y_{b,\tau-l} + \gamma_3^B Z_{\tau-l} + \alpha_b + \varepsilon_{b\tau} \tag{6}$$

where μ^L, μ^B are the intercept terms, X_a, X_b are account-dependent, time-independent covariates, i.e. application variables; $Y_{a,\tau-l}, Y_{b,\tau-l}$ are account-dependent, time-dependent covariates, i.e. behavioural variables, lagged *l* months; $Z_{\tau-l}$ are account-independent, time-dependent covariates, i.e. macroeconomic variables, lagged *l* months; $\gamma_1, \gamma_2, \gamma_3$ are unknown vectors of parameters to be estimated; and $\alpha_a + \varepsilon_{a\tau}, \alpha_b + \varepsilon_{b\tau}$ are the error terms, with $\alpha_a, \alpha_b \sim IID\left(0, \sigma_\alpha{}^2\right)$ and $\varepsilon_{a\tau}, \varepsilon_{b\tau} \sim IID\left(0, \sigma_\varepsilon{}^2\right)$.

The Mixture model could then be used to predict balance at any given time during the loan. This is done by first applying the survival model to all accounts to predict the probability of being overstretched at each duration time τ. Then, regardless of the estimated probability, one applies the balance panel model and the limit panel model onto all observations of all accounts to get an estimated balance and estimated limit, again at each time [4] τ. Because the models would be estimated for the subsets described above, these predicted values, $\hat{B}_{i\tau}$ and $\hat{L}_{i\tau}$, are the values of $B_{i\tau}$ given $B_{b\tau} < L_{b\tau}$ and $L_{i\tau}$ given $B_{a\tau} \geq L_{a\tau}$ respectively, in both cases given default. The final predicted value for balance of an account *i* at duration time τ, given default, $\tilde{B}_{i\tau} \mid d_i = 1$, is then a combination of the repeated events survival model estimating the probability of balance exceeding limit at time τ, and the panel models estimating either balance or limit at time τ. This is the expected value of balance and limit, given the probabilities of the balance exceeding the limit at time τ, and the assumed approximation, as defined in Equation 7 (which is just Equation 2 rewritten in a more efficient form):

$$\left[\tilde{B}_{i\tau} \mid d_i = 1\right] = \left(P(S_{i\tau}) \times \left(\hat{L}_{i\tau} \mid d_i = 1\right)\right) + \left(\left(1 - P(S_{i\tau})\right) \times \left(\hat{B}_{i\tau} \mid d_i = 1\right)\right) \quad , \tag{7}$$

where $P(S_{i\tau}) = P(B_{i\tau} \geq L_{i\tau} \mid d_i = 1)$ and is the estimated probability that account *i* is overstretched at time τ, i.e. that the balance for account *i* at time τ exceeds the limit for account *i* at time τ; and $\hat{L}_{i\tau}$ and $\hat{B}_{i\tau}$ are the estimated values for limit and balance respectively, from their respective panel models.

[4] When predicting balance and limit we set the random effect term at its mean (zero) in every case since its value is unknown for every case that is not in the training sample.

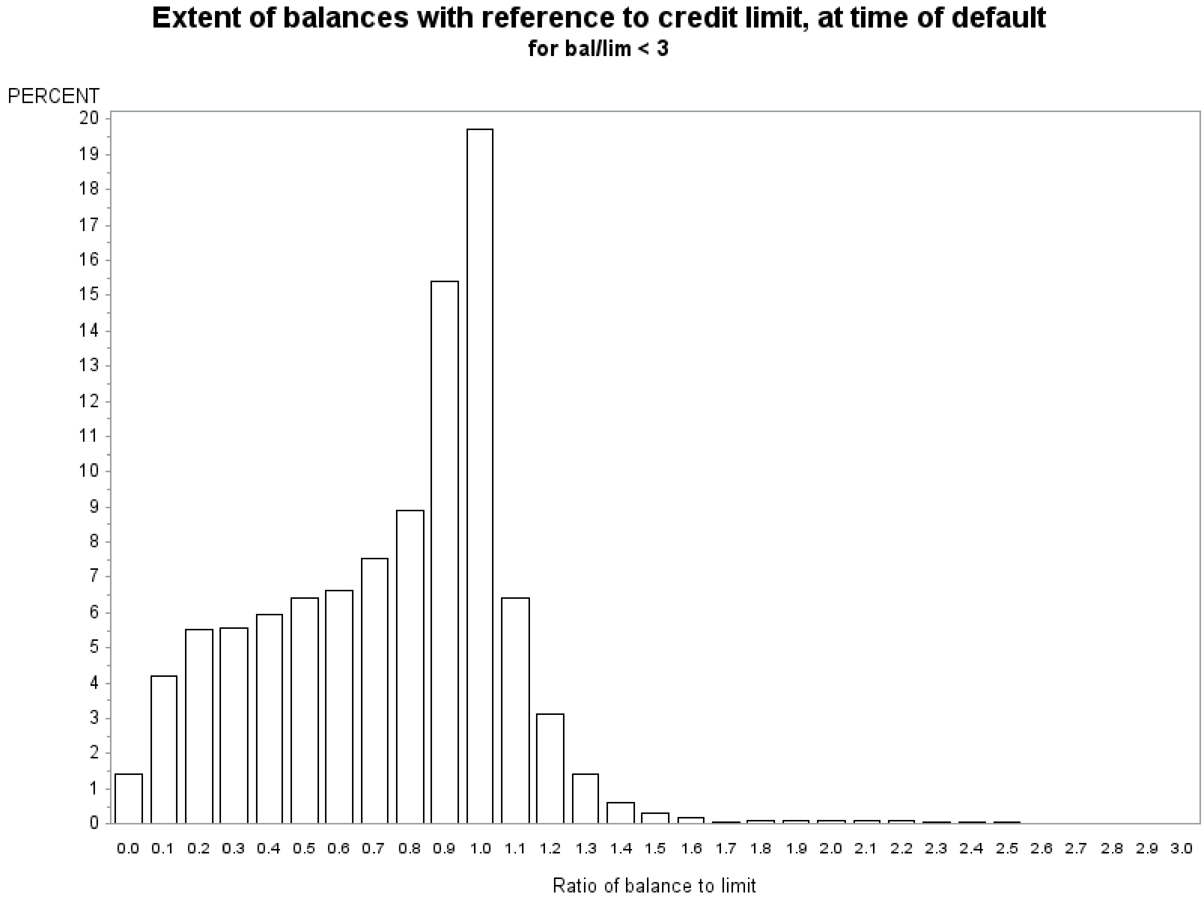
4. Data and variables

4.1. Data

Data is supplied by a major UK bank and consists of a large sample of credit card accounts, geographically representative of the UK market. The accounts were drawn from a single product, and opened between 2001 and 2010. Accounts were observed and tracked monthly up to March 2011 or until it was closed, whichever is earlier. A minimum repayment amount is calculated in each month for each account and accounts progress through states of arrears depending on whether they are able to make the minimum repayment amount. We set the minimum repayment amount at 2.5% of the previous month's outstanding balance or £5, whichever is higher, unless the account is in credit, in which case the minimum repayment amount is £0, or the account has an outstanding balance of less than £5, in which case the minimum repayment amount would be the full outstanding amount. It is also possible for accounts to recover from states of arrears should the borrower make repayment amounts large enough to cover accumulated minimum repayment amounts that were previously missed. An account is then said to go into default if it goes into 3 months in arrears (not necessarily consecutive). For more details on the movement of accounts between states, see Leow and Crook (2014), but note that the percentage used here is different.

Accounts that have a credit limit of £0 at any point in the loan are removed, based on the assumption that these accounts would have been singled out as problem loans by the bank. It is possible for accounts to be in credit, such that balance is negative, so balance is constrained such that observations that have negative balance have £0 balance. We experimented with various lags on the time-dependent covariates in all of the models and report results for lags of 12 and of 6 months. Because of these lags, and the minimum time required for accounts to go into default, we also removed accounts that have been on the books less than 15 and 9 months respectively.

Figure 2: Distribution of ratio of balance over limit at time of default (for ratios less than 3)



From the data, we see that some accounts go into default with an outstanding balance greater than their credit limit. This is illustrated in Figure 2, which gives the distribution of the ratio of balance over limit at the time of default (only for ratios less than 3 for a clearer picture of the distribution). The peak in the graph corresponds to borrowers defaulting with a balance equal to their credit limit, but we also do see a sizeable proportion of borrowers who default with balances on either side of their credit limits.

4.2. Explanatory and macroeconomic variables

Common application variables are available, including age, time at address, time with bank, income, presence of landline and employment type. Behavioural variables are also available on a monthly basis, including repayment amount, credit limit, outstanding balance and number and value of cash withdrawals or card transactions. From these, further behavioural indicators can be derived, for example, the number of times an account oscillates between states of arrears and being up-to-date, the proportion of time the account has been in arrears and the average card transaction value. Any behavioural variables used in the model are lagged 12 (or 6) months.

The macroeconomic variables considered here are listed in Table 2. The main source of macroeconomic variables is the Office of National Statistics (ONS), supplemented by data from Bank of England (BOE), Nationwide and the European Commission (EC) where appropriate. We use the non-seasonally adjusted series unless unavailable because the balance and limit data are also not seasonally adjusted. Any macroeconomic variables used in the model are also lagged 12 (6) months.

Table 2: Description of macroeconomic variables

Variable	**Source (id)**	**Description**
AWEN	ONS (KA5Q)	Average earnings index, including bonus, including arrears, whole economy, not seasonally adjusted
CIRN	BOE (CFMHSDG)	Monthly average of UK resident monetary financial institutions (excl Central Bank) sterling weighted average interest rate , credit card loans to households (%) not seasonally adjusted
CLMN	ONS (BCJB)	Claimant count rate, UK, percentage, not seasonally adjusted
CONS	EC (CONS.UK.TOT.COF.BS.M)	Total consumer confidence indicator, UK, seasonally adjusted
HPIS	Nationwide	House price index ~~A~~all houses, seasonally adjusted
IOPN	ONS (K24V)	Index of production, all production industries, not seasonally adjusted
IRMA	BOE	Monthly average of Bank of England's base rate
LAMN	ONS (BE1)	Log (base e) of total consumer credit, amounts outstanding, not seasonally adjusted
LFTN	ONS	Log (base e) of FTSE all share price index, month end, not seasonally adjusted 10/4/62=100
RPIN	ONS (CHAW)	All items retail price index, not seasonally adjusted, January 1987=100

UERS	ONS (YCNO)	Labour Force Survey unemployment rate, UK, all, ages 16 and over, percentages, seasonally adjusted

ONS denotes Office of National Statistics. BOE denotes Bank of England. Nationwide is Nationwide Building Society. 'id' denotes the data source's identifier for the variable.

4.3. Training and test set split

Although we are interested in the prediction of outstanding balance of an account in each time step, these predictions of balance only become EAD values if and when accounts go into default. We also believe that balances of defaulted and non-defaulted accounts behave differently, and we see from Figure 3 that balances of non-default accounts are on average lower, and have more occurrences of 0 than the balances of default accounts. As such, we only use accounts that do (eventually) go into default. Because we only use observations from accounts that do go into default for the development of the EAD model, we do not need to be concerned with accounts that are inactive, e.g. have zero transactions and zero balance on the card for an extended period of time, but remain in the portfolio.

In the Introduction we explained that the mixture model will both predict balance at each time in the history of a defaulted account as well as at the time of default. The former is useful because a lender does not know when, or if, an account will default. We compare the performance of the established and mixture model in these two settings by using two different test sets as follows. The dataset is divided to give the training set consisting of all accounts that do go into default at some time in their history and were opened on or before 31 December 2008, giving about 94,000 unique accounts. Test set I is an out-of-sample test set and is created using the remaining default accounts, consisting of all observations of all accounts opened on or after 01 January 2009. Test set I consists of about 12,000 unique accounts, giving more than 66,000 month-account observations. Test set II is created as a subset of Test set I, where only observations at the time of default are included. Test set I would give an indication of how well the model is able to predict balance for accounts that are likely to be delinquent but may not yet have gone into default at each time in their account history, whilst Test set II would be an indication of how well the model is able to predict at default-time, regulatory EAD. The relationship of the training and test sets are represented in Figure 1. We calculate several performance measures including ~~r-squared~~R-squared values, for the two test sets: Test set I, for all accounts, for all observation times; and Test set II, for all accounts, only at time of default.

The portfolio of non-default accounts is not used in either the modelling or the testing as we estimate balance given default. Applying the Mixture model to observations of non-default accounts would give us the predicted balance should the account go into default, which is different to the observed balance, as seen in Figure 3, which would mean that we will not be able to score how well the model is predicting.

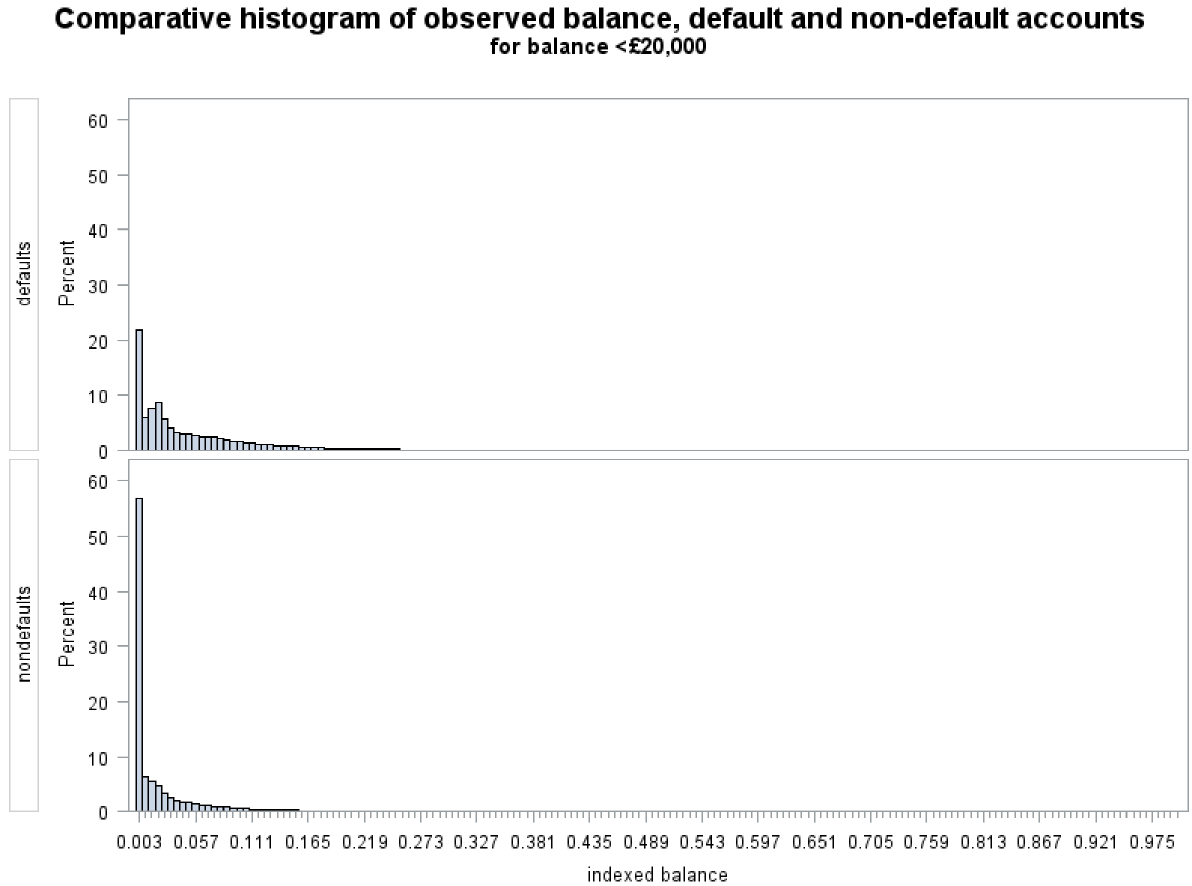


Figure 3: Distributions of observed balance, for default and non-default accounts, for balance less than £20,000.

4.4. Model estimation

Both panel models were estimated using Generalised Least Squares (GLS) estimators. We estimated models with lags of 12 months and of 6 months lags. Covariates include application variables, lagged behavioural variables, and lagged macroeconomic variables, defined in Equations 5 and 6. We initially estimated the survival model and models for balance and limit, separately, using a very large number of application, behavioural and macroeconomic variables with 12 month lags. Covariates were then retained or deleted based on their level of statistical significance including that of other variables, their relevance and the predictive accuracy of the overall model. So for example in the limit and balance equations Time at address (TAAdd) and a binary variable indicator for missing or unknown time with bank (TWBank_MU) were not significant in the balance and limit equations and so were not included in the final equation (see the Appendix) whereas they were significant in the survival model and so were included in that[5]. Thus different sets of parameters are used in each model and between the lagged models.

The survival model did not include utilisation or credit limit because although they were very statistically significant, the overall accuracy of the model at lag 12 months was slightly lower when the combination of variables that included these two was used. At lag 6 months, inclusion or exclusion of these two variables actually made little difference to predictive accuracy. We found the greatest predictive accuracy was gained when the training set of the balance model was restricted to cases when the minimum balance was over £200. Since each account typically has multiple observations (month-account observations), we adjusted for serial correlation by using a clustered sandwich estimator (on account ID) to estimate variance and standard errors Drukker (2003).

To compare the predictive accuracy of the Mixture model with established methods, we use the training set with observations only at time of default to estimate the EADF, LEQ and CCF cross-sectional regression models (represented by the dotted square from the training set in Figure 1). For all observations at time of default, EADF, CCF and LEQ are predicted based on observed covariates lagged 12 (6) months before default, according to the equations in Table

[5] The omission of lagged utilisation in the limit and balance equations allows more flexibility in the estimated parameters concerning lagged balance and limit.

1. Similar to the EAD papers mentioned in the literature, some observations were further excluded from this subset due to some very extreme observations of CCF and LEQ. The final number of accounts and observations used in each training set when covariates were lagged 12 months is given in Table 3.

Table 3: Number of observations for balance and limit subsets, lag 12 months

Model	**Number of accounts**	**Number of observations**	**Minimum observations for any account**	**Maximum observations for any account**	**Average observations per account**
Balance	13,859	184,608	4	105	13.3
Limit	36,453	798,486	4	107	21.9
CCF	43,686	43,686	1	1	1
EADF	68,47~~6~~9	68,479	1	1	1
LEQ	31,821	31,821	1	1	1

EADF, LEQ and CCF were regressed on the same covariates as those used in the survival model[6]. A variety of experimentation in terms of modelling functions and techniques was done for these competing models to improve the predictions for these variables. For the modelling of EADF, we tried several functional forms including a beta function and logit link functions but found that a liner model with OLS estimators gave the greatest predictive accuracy. For LEQ, various values of outliers were deleted but the greatest predictive accuracy was gained when we took only values in the range 0 < LEQ < 1 and adopted a generalised linear model with a logit link function with a maximum likelihood estimator. For the CCF model we took a $\log_e$ transformation to transform the distribution to be close to normal, then deleted various sizes of outliers and used an OLS estimator. The predictive accuracy was very poor until we deleted all observations above the 80^{th} percentile. These models are then applied onto the test sets (Test set I and Test set II) and performance measures are calculated. These three regression models are not further documented in this paper.

[6] Except for time varying duration time since last event, duration time squared and number of times event has happened which are all survival model specific and time on books that was included in the competing models but not the survival model.

5. Results

5.1. Survival model for being overstretched

The parameter estimates for the discrete-time repeated events survival model predicting for the event overstretched is given in the appendix, Table A2. We find that the signs of the parameter estimates are intuitive: for example, the probability of being overstretched decreases with age as well as with higher income. In terms of behavioural variables, we find that the probability of being overstretched reflects how well borrowers manage their accounts, so borrowers who move in and out of arrears frequently (see rate of total jumps) or are frequently in arrears (see proportion of months in arrears) tend to have a higher probability of being overstretched. In terms of macroeconomic variables, an increase in housing or financial wealth, for example, an increase in the House Price Index (HPI) would decrease the probability of being overstretched; but easier access to credit (indicated by an increase in credit amount outstanding) increases the probability of being overstretched.

5.2. Panel models for balance and limit

The parameter estimates for both panel models are given in Table A2 in the appendix. We acknowledge that the balance from 12 months previous is included as a variable in the balance model, and credit limit from 12 months previous is included as a variable in the limit model. Although this would raise the issue of endogeneity in econometric interpretation, it is not an issue in this case as we are using the model solely for the purpose of prediction. Although the panel models are developed with random effects, these random effects are not known for accounts in the test set(s). The random effects associated with each account in the test set is assigned to be the mean values of α_i and ε_{it}, that is zero in both cases.

The goodness of fit statistics for the panel models for balance and limit, based on the training set with time varying covariates lagged 12 months are given in the appendix, Table A1. We expect it to be easier to predict the limit, as this would be based on a combination of application time and behavioural indicators, and is reflected in the impressive ~~r-squared~~R-squared value for the limit model. The panel model for balance does not predict as well as that for the limit, as factors affecting outstanding balance of an account would include borrower circumstances which would be impossible to take into account given the information we have.

5.3. Overall performance

After applying the Mixture model onto the test sets, we compute overall ~~r-squared~~R-squared, Mean Absolute Error (MAE), Mean Error (ME) and the symmetric Mean Absolute Percentage Error (sMAPE) for the predicted versus the observed balance, i.e. we transform the predicted CCF, LEQ and EADF into predicted balances, given in Table 4. The sMAPE is able to circumvent the problem of having £0 balance that would mean dividing by 0 in the calculation of MAPE.

Table 4: Performance measures for Mixture, LEQ, EADF and CCF models, for test sets, based on predicted balances

Model	**Test set**	**Lag 12**					**Lag 6**				
		No.of obs	**R-squared**	**MAE**	**ME:Obs-Pred**	**sMAPE**	**No. of obs**	**R-squared**	**MAE**	**ME:Obs-Pred**	**sMAPE**
Mixture model developed on default accounts, min balance >£200	Test set I	18,584	0.5565	646.20	-125.98~~21.36~~	0.5277	66,460	0.5814	652.26	-187.48~~10.65~~	0.5432
	Test set II	4,122	0.6321	611.54	21.36~~-125.98~~	0.4369	11,734	0.6564	647.61	-10.65~~187.48~~	0.3923
LEQ model developed on default accounts at time of default, 0 < LEQ≤~~<~~1	Test set I	18,584	0.4928	632.48	157.82	0.4900	66,460	0.4790	634.90	-263.02	0.4510
	Test set II	4,122	0.5673	632.37	292.53	0.4114	11,734	0.6272	638.83	-154.60	0.3109
EADF model developed on default accounts at time of default	Test set I	18,584	0.5360	613.01	-218.80~~-35.06~~	0.3956	66,460	0.3903	692.46	-316.58	0.448~~0~~4
	Test set II	4,122	0.6981	548.63	-35.02~~-218.80~~	0.2977	11,734	0.5789	675.99	-127.62	0.3012
CCF model (ln CCF) developed on default	Test set I	18.584	-0.0009	962.94	730.97	1.1890	66,460	0.2275	805.27	572.45	0.9463
	Test set II	4,122	0.0492	1024.80	807.10	1.1183	11,734	0.1975	931.02	732.56	0.8282

accounts at time to default, CCF>0 and truncated at 80th percentile

We omit cases where both observed and predicted balance are £0 (i.e. the prediction is accurate and there is 0 error) from the calculation of sMAPE as they do not contribute to the error. We see that when considering predictions 12 months in advance (left hand panel) the Mixture model is able to achieve an ~~r-squared~~R-squared of 0.56 when predicting for balances for accounts that are likely to be delinquent, at all times they are observed. This is an improvement from the ~~r-squared~~R-squared values of between 0.54 for EADF, 0.49 for LEQ and -0.001 for CCF[7] . The Mixture model also has the lowest ME. But in terms of MAE and sMAPE the EADF method gives lower errors. When considering balance at the time of default, the Mixture model has a lower ME at £21.36~~-£126~~ than the EADF (-£35.02~~-£219~~) and LEQ (£293); in terms of ~~r-squared~~R-squared and MAE its performance is inferior to the EADF though better than the other two methods, and in terms of sMAPE, its performance is below those of EADF and LEQ.

The Mixture model gives a prediction at each duration time since the opening of the account. When we consider the performance at a prediction horizon of, say, 6 months (right hand panel) we see that for accounts at all observation times, the ~~r-squared~~R-squared of the Mixture model at 0.58 is considerably above those of the other methods, the largest of which is LEQ at 0.48 with EADF at 0.39. In terms of ME, the Mixture model is also considerably more accurate than the other methods, with a ME of ~~£11~~£187.48 whilst the closest of the other methods is -£263 for LEQ. In terms of MAE, the Mixture model is more accurate than EADF but less so than LEQ. At the time of default, the Mixture model has the highest ~~r-squared~~R-squared at 0.66 ~~although~~and it is far more accurate than the other methods in terms of mean error with a mean error of just £10.65 compared with that of EADF of -£127.62. It is more accurate in terms of MAE as well, although less accurate on sMAPE.~~less accurate than the LEQ and EADF methods in terms of the error metrics.~~

[7] ~~r-squared~~R-squared is computed as 1-(sum of squared errors/total sum of squares). The predicted values are values of EAD predicted by the relevant model and the observed values are the values observed in the data. R-squared can be negative when predicted and observed values are compared and the implied model does not have a constant as is the case when predicting balance from the CCF model.

It is difficult to compare our results with those of the literature because many other studies quote only statistics relating to the regression model and not for values of predicted EAD. Thus the regression models developed for credit cards LEQ by Qi (2009) achieved adjusted ~~r-squared~~R-squared values of between 0.06 to 0.37, on a sample of default time observations depending on whether the accounts were current or delinquent, and whether outliers were excluded from the model development. Jacobs Jr. (2008), working on corporate data, achieved pseudo median ~~r-squared~~R-squared values of 0.15, 0.19 and 0.13 for LEQ, CCF and EADF respectively. Barakova and Parthasarathy (2013) find adjusted ~~r-squared~~R-squared values for different models for corporate loans of between 1% and 33% depending on the model and treatment of outliers. In contrast, Yang and Tkachenko quote an ~~r-squared~~R-squared of 0.91 for EAD using EADF with a least squares logit algorithm but that is for a sample of corporate borrowers and we do not know if this applies to a testing sample.

Figure 4 compares the distributions of predicted and observed balances for Test set II, i.e. only default time observations for all default accounts. The values of balance are limited to between £0 and £20,000 for clearer representation of the distributions and all values of balances are indexed on some value of observed balance. The Mixture model predicts the mean with considerable accuracy (a difference in indexed value of +0.0003) compared with the EADF and LEQ models (with differences of +0.0027 and -0.0156 respectively). The CCF is again the least accurate by a large margin.

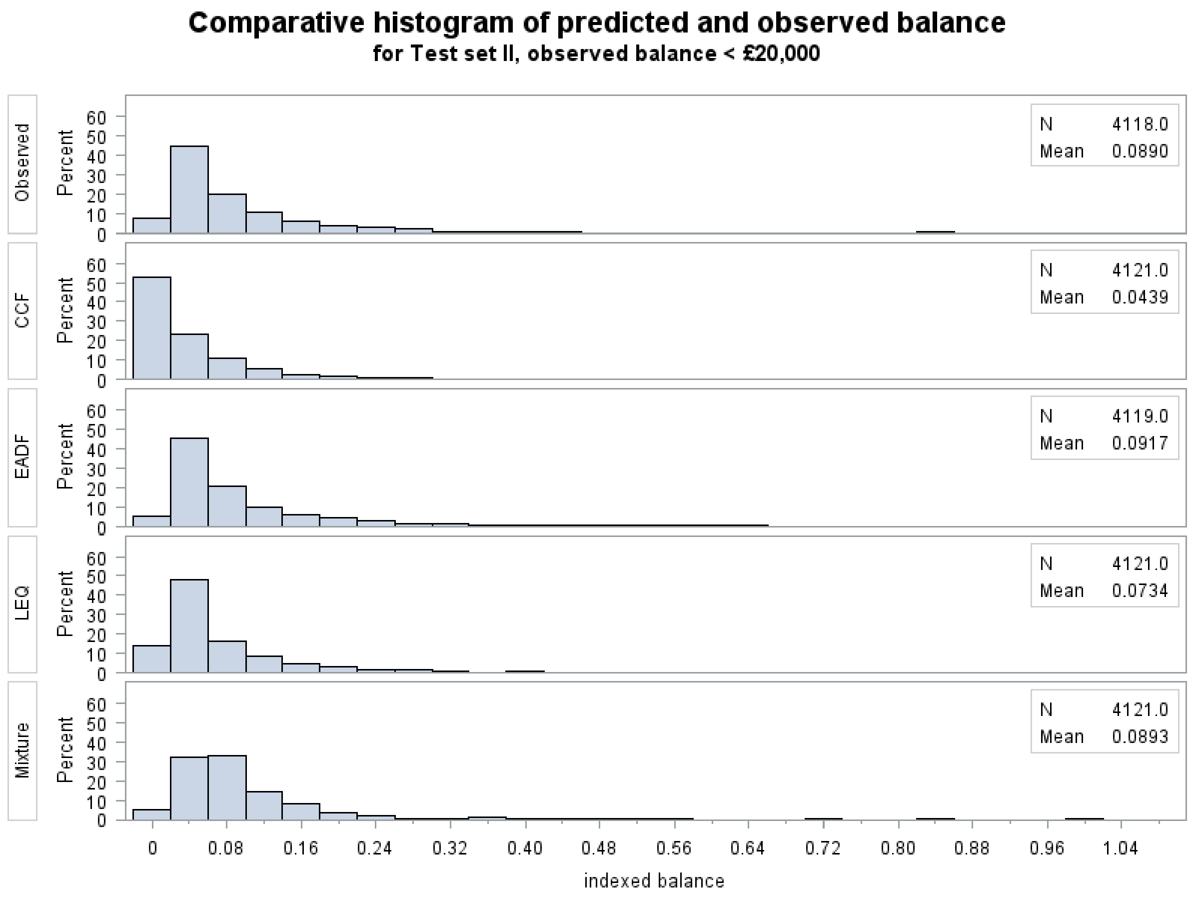


Figure 4: Comparative histogram of predicted and observed balances, indexed on observed balance, for Test set II, only observations at time of default (where observed balance lies between £0 and £20,000).

It is interesting to note that, although with variables lagged 12 months, the EADF model has a higher ~~r-squared~~R-squared and lower MAE and sMAPE than the Mixture model at default time (Table 4), when we plot the distributions (Figure 4), the Mixture model yields more accurate predictions compared to the EADF model in terms of the mean. This suggests that whilst the MAE value for the Mixture model shows the deviations from the observed values are, on average, larger for the Mixture than for the EADF model at default time, the net value is closer to the observed value for the Mixture than for the EADF model. Looking at the distributions, the Mixture model is less accurate than EADF for the smaller values of balance but more accurate for the larger values. Arguably, the larger values are the balances that a portfolio manager would most be concerned about. Overall then, we believe that the Mixture model is a more accurate and useful model to use to predict EAD and outstanding balances for accounts likely to default at pre-default times than are conventional models.

6. Concluding Remarks

We propose a Mixture model to predict for credit card balance at any time τ, given that an account has defaulted. We exploit the advantage that this model has over conventional cross-section models of incorporating the movement in balance and in limit over time as the account moves towards default. Specifically the method involves first estimating a discrete-time repeated events survival model to estimate the probability of an account being overstretched, i.e. having a balance greater than its limit, at any time τ. Next, two panel models with random effects are developed to estimate balance and limit separately, at any time τ. The final prediction for balance at duration time τ is then taken as the sum of two products, all at time τ: the probability of being overstretched multiplied by the estimated limit; and the probability of not being overstretched multiplied by the estimated balance in both cases given default (c.f. Equation 7).

Applying this Mixture model to a large portfolio of default loans and their historical observations, we find that we are able to get good predictions for outstanding balance for accounts that at some time default, not only at the time of default, but at any time over their entire loan period. This would allow us to make predictions for outstanding balance and hence EAD before default occurs, for delinquent accounts. Considering predictions 12 months into the future ~~We~~we find that at the time of default, the EADF model gives results that are, on three ~~some~~ measures, more accurate and on one measure~~others~~ less accurate than the Mixture model. However the Mixture model is more accurate in terms of the mean and mean error and has the added advantage of giving more accurate predictions for larger balances than EADF. Turning to predictions before the time of default, the Mixture model has the highest ~~r-squared~~R-squared and smallest mean error of any of the methods. If one wishes predictions a mere 6 months into the future the Mixture model has the highest ~~r-squared~~R-squared at both default time and at earlier times and the lowest mean error by a considerable margin. However, ~~Overall,~~ whilst ~~we believe~~ the Mixture model is a competitive methodology ~~better methodological choice~~ for the prediction of balance for accounts that are likely to default, especially if a prediction 6 months into the future is required, further research is desirable to explore its accuracy in other datasets.

It is appropriate to remark that some types of portfolios, such as corporate portfolios will differ in the proportion of cases where the balance at default exceeds the limit and so the sampling variance of the estimated parameters of the survival model would differ between portfolios.

Following this work, we plan to incorporate stress testing into our risk models. We plan to combine PD, LGD and EAD models, and to stress test each component model independently yet retain the knock-on effects in an adverse economic situation, if any. The obvious covariates to stress test within the models would be the macroeconomic variables; however, we would also like to consider methods which would allow us to stress the behavioural variables as well. It is not always clear how behavioural variables are affected by the economy, especially in the case of retail loans where the economy is expected to affect individuals differently and to varying degrees. The different combinations of $PD_{i\tau}$, $LGD_{i\tau}$ and $EAD_{i\tau}$ computed would enable us to get a distribution for $loss_{i\tau}$, from which we expect to be able to predict for expected and unexpected losses better.

References

ARATEN, M. & JACOBS, M. J. 2001. Loan Equivalents for Revolving Credits and Advised Lines. *The RMA Journal*.

BARAKOVA, I. & PARTHASARATHY, H. 2013. Modeling Corporate Exposure at Default. *Office of the Comptroller of the Currency Working Paper*.

BASEL COMMITTEE ON BANKING SUPERVISION 2004. International Convergence of Capital Measurement and Capital Standards: A Revised Framework.

CAMERON, A. C. & TRIVEDI, P. K. 2005. *Microeconometrics: Methods and Applications*, Cambridge University Press.

DRUKKER, D. M. 2003. Testing for Serial Correlation in Linear Panel-Data Models. *The Stata Journal,* 3**,** 168-177.

GUJARATI, D. N. 2003. *Basic Econometrics*, McGraw Hill.

JACOBS JR., M. 2008. An Empirical Study of Exposure at Default. *Office of the Comptroller of the Currency Working Paper*.

JIMÉNEZ, G., LOPEZ, J. A. & SAURINA, J. 2009. Empirical Analysis of Corporate Credit Lines. *Review of Financial Studies,* 22**,** 5069-5098.

JIMÉNEZ, G. & MENCÍA, J. 2009. Modelling the distribution of credit losses with observable and latent factors. *Journal of Empirical Finance,* 16**,** 235-253.

LEOW, M. & CROOK, J. N. 2014. Intensity Models and Transition Probabilities for Credit Card Loan Delinquencies. *European Journal of Operational Research,* 236**,** 685-694.

MORAL, G. 2006. EAD Estimates for Facilities with Explicit Limits. *In:* ENGELMANN, B. & RAUHMEIER, R. (eds.) *The Basel II Risk Parameters.* Springer Berlin Heidelberg.

QI, M. 2009. Exposure at Default of Unsecured Credit Cards. *Office of the Comptroller of the Currency Working Paper*.

RISK MANAGEMENT ASSOCIATION 2004. Industry Practices in Estimating EAD and LGD for Revolving Consumer Credits - Cards and Home Equity Lines of Credit.

TAPLIN, R., TO, H. M. & HEE, J. 2007. Modelling Exposure at Default, Credit Conversion Factors and the Basel II Accord. *Journal of Credit Risk,* 3**,** 75-84.

THOMAS, L. C. 2010. Consumer Finance: Challenges for Operational Research. *Journal of the Operational Research Society,* 61**,** 41-52.

YANG, B. H. & TKACHENKO, M. 2012. Modelling Exposure at Default and Loss Given Default: Empirical Approaches and Technical Implementation. *Journal of Credit Risk*, 8(2), 81-101.

VERBEEK, M. 2004. *A Guide to Modern Econometrics*, John Wiley & Sons.

APPENDIX

Table A1: Performance indicators for panel models, for training set

Model	Overall R-squared (train)	σ_u	σ_e	ρ
Balance	0.4673	866.3302	828.6274	0.5159
Limit	0.9005	403.5576	482.6812	0.4114

Table A2: Parameter estimates of survival model for event overstretched and panel models for balance and limit lag 12

Code	Parameter	Discrete-time repeated events survival model for P(B>=L)			Panel model with random effects for balance			Panel model with random effects for limit		
		Estimate	WaldChiSq	ProbChiSq	Estimate	z	P>|z|	Estimate	z	P>|z|
Intercept	Intercept	-9.7726	20.1402	<.0001	-1,752.36	-3.40	0.001	37,574.49	37.11	<.0001
Application variables										
ageapp_1	Age at application group 1	-	-	-	-	-	-	-	-	-
ageapp_2	Age at application group 2	-0.1447	43.6769	<.0001	-39.6967	-1.43	0.152	24.2148	2.91	0.004
ageapp_3	Age at application group 3	-0.1963	64.0319	<.0001	-2.0937	-0.06	0.950	51.2779	4.65	<.0001
ageapp_4	Age at application group 4	-0.1413	25.8411	.<0001	11.4810	0.32	0.750	82.4971	6.10	<.0001
ageapp_5	Age at application group 5	-0.1338	19.9317	<.0001	24.6703	0.63	0.531	137.193	8.37	<.0001
ageapp_6	Age at application group 6	-0.2033	39.0898	<.0001	37.2720	0.89	0.374	158.7112	8.86	<.0001
ageapp_7	Age at application group 7	-0.2078	31.5887	<.0001	81.8585	1.89	0.059	178.8056	8.24	<.0001
ageapp_8	Age at application group 8	-0.3006	42.2476	<.0001	88.9510	1.90	0.058	186.5624	7.57	<.0001
ageapp_9	Age at application group 9	-0.4161	51.8056	<.0001	155.425	2.84	0.005	172.7288	4.88	<.0001

ageapp_10	Age at application group 10	-0.5906	81.2618	<.0001	92.9561	1.69	0.090	127.999	3.78	<.0001
ECode_A	Employment code, group A	-	-	-	-	-	-	-	-	-
ECode_B	Employment code, group B	-0.0256	1.1698	0.2794	63.4967	1.68	0.093	-12.8937	-0.84	0.401
ECode_C	Employment code, group C	0.0962	2.3881	0.1223	-23.3239	-0.48	0.629	-12.7538	0.47	0.638
ECode_D	Employment code, group D	-0.2193	49.4397	<.0001	-69.9821	-2.16	0.031	244.4278	20.87	<.0001
ECode_E	Employment code, group E	-0.1671	81.3304	<.0001	-23.8745	-0.90	0.367	222.8055	18.53	<.0001
INC_L	Income, ln	-0.1651	199.9145	<.0001	329.721	7.95	<.0001	389.5459	25.57	<.0001
INC_M0	Binary indicator for missing or 0 income	-1.5540	196.3966	<.0001	2927.806	7.72	<.0001	3421.209	24.18	<.0001
LLine	Binary indicator for presence of landline	0.0084	0.1865	0.6659	86.6014	3.28	<.0001	-	-	-
NOCards	Number of cards	-0.0664	92.6344	<.0001	62.5053	4.86	<.0001	51.8571	10.34	<.0001
TAAdd	Time at address (years)	0.0008	0.6132	0.4336	-	-	-	-	-	-
TWBank_MU	Binary indicator for missing or unknown time with bank	-0.0792	11.6248	0.0007	-	-	-	-	-	-
TWBank	Time with bank (years)	-0.0014	211.2680	<.0001	-0.2347	-2.26	0.024	1.0901	17.73	<.0001
X_A	Variable X, group A	-	-	-	-	-	-	-	-	-
X_B	Variable X, group B	0.3040	207.3578	<.0001	-9.3027	-0.32	0.751	--226.1428	-18.05	<.0001
X_C	Variable X, group C	0.3935	222.4255	<.0001	-71.7786	-2.34	0.019	-205.4936	-15.37	<.0001
X_D	Variable X, group D	0.2797	129.8488	<.0001	-26.8704	-0.98	0.326	-154.2192	-12.29	<.0001
X_E	Variable X, group E	0.5082	478.2932	<.0001	-9.7506	-0.29	0.772	-499.7334	-31.27	<.0001
Behavioural variables, lagged 12 months										
ATRV_lag12	Average transaction value	-0.0008	230.0800	<.0001	0.1461	8.74	<.0001	0.0491	3.40	0.001
CASC_lag12	Number of cash withdrawals	0.0953	36.1581	<.0001	-	-	-	-	-	-
CASV_lag12	Amount of cash withdrawal	0.0001	4.3231	0.0376	-	-	-	-	-	-

CRLM_lag12	Credit limit	-	-	-	0.2910	13.67	<.0001	0.6948	59.02	<.0001
JUMP_lag12	Rate of total jumps	0.1753	9.0396	0.0026	403.2316	6.96	<.0001	-97.3752	-4.23	<.0001
PARR_lag12	Proportion of months in arrears	0.2543	12.7681	0.0004	-404.4825	-6.38	<.0001	112.979	3.92	<.0001
PAYM_lag12	Repayment amount	-0.0001	30.2486-	<.0001	-	-	-	0.0283	6.27	<.0001
SCBA_lag12	Outstanding balance	-	-	-	0.1012	9.45	<.0001	0.0386	2.18	0.029
Macroeconomic variables, lagged 12 months										
AWEN_lag12	Average wage earnings	-.0.0020	6.6668	0.0098	-0.1042	-0.15	0.879	-1.4374	-9.60	<.0001
CIRN_lag12	Credit card interest rate	0.1180	112.4658	<.0001	80.2595	6.48	<.0001	-116.2638	-32.44	<.0001
CONS_lag12	Consumer confidence	0.0075	35.6671	<.0001	5.4994	3.74	<.0001	-	-	-
HPIS_lag12	House Price Index	-0.0024	22.7570	<.0001	-1.5996	-2.83	0.005	6.8107	45.93	<.0001
IOPN_lag12	Index of production	-0.0008	2.5745	0.1086	-	-	-	-	-	-
IRMA_lag12	Base interest rate	-0.1107	93.8650	<.0001	--51.4505	-3.39	0.001	-125.5614	-27.29	<.0001
LAMN_lag12	Amount outstanding, ln	0.8548	20.3212	<.0001	-	-	-	-2996.797	-34.24	<.0001
LFTN_lag12	FTSE Index, ln		-	-	-169.2373	-3.21	0.001	-686.573	39.86	<.0001
RPIN_lag12	Retail Price Index	-	-	-	-0.4331	-0.16	0.875	13.3002	15.75	<.0001
UERS_lag12	Unemployment rate	-0.1843	58.6602	<.0001	-54.1236	10.39	<.0001	-219.8914	-24.13	<.0001
Model specific required variables										
duration	Survival time (months) since last event	-0.0472	1433.3820	<.0001	-	-	-	-	-	-
durnsq	Survival time since last event squared	0.0002	78.5404	<.0001						
period	Number of times event has happened	0.2005	1178.3776	<.0001	-	-	-	-	-	-
Time on books	Time on books (months)	-	-	-	5.4430	10.33	<.0001	9.5058	37.11	<.0001